Facets of native-likeness

Bergmann, Christopher

5

The effect of formulaic language on speech fluency

This paper is in preparation:
Sprenger, S. A., Bergmann, C., & Schmid, M. S. (in preparation).
The effect of formulaic language on speech fluency among monolingual and bilingual speakers of German.

Abstract
The use of formulaic language has been linked to speech fluency, in particular for bilingual speakers, as the direct retrieval of word sequences from memory is hypothesized to ease the speaker's processing burden. However, second-language (L2) speakers typically have a smaller repertoire and possibly less ready accessibility of these sequences. These factors may have a negative impact on their speech fluency (Pawley & Syder, 1983). Here, we tested these hypotheses in samples of spontaneous speech produced by three groups of speakers of German: monolingual speakers, tested in Germany; first-language (L1) attriters (i.e., bilingual native speakers of German, immersed in their L2, English), tested in the United States of America; and L2 speakers (L1: English), tested in Germany. Our analysis of their use of formulaic language showed on the one hand that lexical collocations (as defined by Granger & Paquot, 2008) were the only type of sequence that L2 speakers used significantly less frequently than the two groups of native speakers. On the other hand, a correlation of fluent speech with the use of lexical collocations was confirmed for L2 speakers and L1 attriters, that is, for both bilingual populations. The use of grammatical collocations, by contrast, was only correlated with global proficiency, not with fluency. These results – demonstrating a link between some types of formulaic sequences and speech fluency, especially in bilingual speakers – strongly support Pawley and Syder's (1983) hypothesis that speech fluency is a function of the use of formulaic language.

5.1 Introduction

In a second language (L2), fluency in spontaneous speech is often strived for, but not easily achieved (Pawley & Syder, 1983). The majority of L2 speakers face a *fluency gap* (Segalowitz, 2010), that is, they are rarely able to use their L2 with the same level of skill as their first language (L1) (e.g., Guz, 2015; Kahng, 2014; Raupach, 1984; Towell, Hawkins, & Bazergui, 1996; Wiese, 1984). In our own study on utterance fluency in monolingual and bilingual speakers of German (Bergmann et al., 2015), we observed more pauses, repetitions and self-corrections in bilingual than in monolingual speakers.

The reasons for this lack of fluency in L2 speech are manifold: Even proficient L2 speakers often have not acquired the language to the full extent. This concerns the representations involved, the rules according to which they need to be combined, and the degree to which retrieval and grammatical encoding are proceduralized (R. Schmidt, 1992). In addition, bilingual speakers are to a considerable degree affected by language competition. Highly active representations and routines from the non-target language interfere with access to those in the target language (Bergmann et al., 2015; Kroll et al., 2008, 2006; Linck, Kroll, & Sunderman, 2009).

A factor that may add to problems with fluency is a lack of familiarity with conventionalized multi-word expressions, such as *How do you do?* In almost any communicative situation, L1 speakers use a rich repertoire of such ready-made phrases, potentially enabling them to skip some of the burdensome processing steps that make L2 speakers stumble (Pawley & Syder, 1983). Here, we focus on this particular aspect of L2 proficiency and its relationship to fluency. Specifically, we will explore to what extent L2 speakers make use of multi-word expressions in their spontaneous speech, whether their use of such items is related to other linguistic measures, and how it relates to their speech fluency. To anchor our observations in native-speaker behavior, we will compare the L2 data to those of monolingual controls and L1 attriters (i.e., L1 speakers who, after emigration, have been immersed in an L2 environment for an extended period). This will allow us to tease apart the aspect of competition on the one hand, which L1 attriters should experience similarly to L2 speakers, and automatization on the other, since attriters should have the same repertoire of such multi-word items to draw upon as monolingual natives.

5.1.1 Theoretical background

In a seminal paper, Pawley and Syder (1983) drew attention to the fact that the ability to use frequently reoccurring phrases and expressions – such as *How do you do?, It is twenty to six,* or *I will look it up for you* – is an important aspect of native proficiency that non-native speakers struggle with. They point out that, in addition to grammatical rules, speakers must learn "a way of distinguishing those usages that are normal or unmarked from those that are unnatural or highly marked" (p. 194). Pawley and Syder refer to this ability as *nativelike selection* and suggest that native speakers must have a large repertoire of familiar expressions that refer to common cultural concepts in a way that is familiar to both speaker and listener. They hypothesize that speakers who have access to such expressions not only come across as more native-like, but are also able to speak more fluently. This fluency benefit is assumed to arise from the advantage of retrieval over computation: Direct retrieval of ready-made sequences will free some of the processing capacity that would otherwise be required for lexical and grammatical formulation processes. These resources can then be used for speech planning or monitoring. From this perspective, being fluent in a language is not only a matter of retrieving individual words and knowing how to combine them, but also depends on our ability to store and access specific instantiations of these combinations in long-term memory.

Over the past thirty years, the idea of memorized sequences as the building blocks of language has gained considerable ground, which has been reflected in the development of syntactic theories that put strong emphasis on the lexicon (e.g., Fillmore, Kay, & O'Connor, 1988; Goldberg, 1995, 2006; Jackendoff, 1997; Langacker, 2008), as well as in usage-based approaches to first- and L2 acquisition (e.g., Smiskova, Verspoor, & Lowie, 2012; Tomasello, 2003). In the domain of L2 acquisition, Segalowitz and Hulstijn (2005) have speculated that L2 speech fluency depends, at least partly, on "the greater availability of ever-larger preassembled linguistic units and the reduced need to compute information" (p. 385). Yet, while there is a considerable number of studies reporting processing advantages for multi-word sequences in the domain of L1 and L2 language comprehension, only few studies have taken a closer look at the relation between formulaic language and speech fluency.

5.1.2 What are multi-word sequences?

Before discussing these studies, however, we turn to the complex question of how to define multi-word sequences as a linguistic unit. The linguistic literature offers a plethora of approaches to the description of larger – and to some extent fixed – expressions.

Maybe due to its broadness and inclusivity, the definition by Wray (2002) is the one that has probably been cited most widely in the field of phraseology: A formulaic sequence is "a sequence, continuous or discontinuous, of words" that is prefabricated, which means "stored and retrieved whole from memory [...] rather than being subject to generation [...] by the language grammar" (p. 9). In this definition, two of three important aspects of formulaic language are summarized: First, it underlines that research in this area centers on how fixed lexical content is expressed through *combinations* of words, rather than through individual lexical items. Second, Wray's definition accentuates the 'non-creative' aspect of formulaic language: Sequences are not formulated on the spot using a grammar that can generate a vast number of utterances, but are drawn holistically from a limited lexicon. The processing advantages that are associated with this type of language and implicit in the definition will be detailed below. A third aspect – less prominent in the present definition – is emphasized in Pawley and Syder's approach: Formulaic language can be seen as "the key to idiomaticity" (Wray, 1999), that is, to native-likeness in social interaction. Summing up, formulaicity can be described as language use characterized by fixedness on many levels: Conventionalized content is expressed by immutable memorized sequences (though there may be some morphological or syntactic freedom; e.g., N. Schmitt & Carter, 2004), the use of which is agreed upon within a community of speakers.

Within these constraints, however, we encounter a variety of different types of formulaic sequences. Every script and schema that we access in a given situation may be represented with its own set of fixed phrases, implying that – in the native language – we talk about a number of representations that equals the number of single words (Becker, 1975; Jackendoff, 1995; Mel'čuk, 1995; Sprenger, 2003). Focusing on different aspects of multi-word expressions, various typologies have been put forward (see Granger & Paquot, 2008, for an overview). A common denominator across categories is that more than one word is used to express a single concept, that these combinations must be learned, and that they are part of the native vocabulary.

In the following, we will refer to these combinations as *formulaic sequences* (Wray, 2002) or, more generally, as *formulaic language,* which will function as an umbrella term for various specific types and will be familiar to readers of the psycholinguistic literature on the processing of larger chunks (e.g., Conklin & Schmitt, 2012; Underwood, Schmitt, & Galpin, 2004). Just like Conklin and Schmitt (2008), we wish the term *formulaic sequence* "to be interpreted as a broad cover term, including all of the various types of multi-word units and collocations, including, but not limited to, idioms, lexical bundles, and lexical phrases" (p. 86). The subtypes (such as idioms or different types of collocations) that we have differentiated in the present study follow the typology by Granger and Paquot (2008) and will be further elaborated on in the methods section.

5.1.3 Evidence for the relation between formulaic language and fluency

From a cognitive processing point of view, Pawley and Syder's (1983) hypothesis that the retrieval of larger chunks of speech may be the key to native-like fluency intuitively makes sense, as – by definition – computation processes are costly, both with respect to time and processing capacity, and bear the risk of errors. Retrieval processes, by contrast, are fast and relatively effortless, at the cost of lower flexibility. This principle also holds for the production of familiar phrases, like standard clock times (Sprenger & Van Rijn, 2013).[4]

Despite the relatively clear theoretical assumptions about the effect of formulaic language on production, there are to date surprisingly few studies that have systematically studied the relationship between the use of formulaic sequences and speech fluency, both in L1 and L2. A well-known and widely cited study on L1 fluency is a qualitative analysis by Kuiper (1996): He analyzed the spontaneous speech of L1 speakers who, due to the context in which they perform, need to maximize their speech fluency. Horse racing commentators, for example, need to be fluent if they want to keep up with the fast-paced sequence of events in a race. At the same time, they must have a lot of information about the horses and their owners, ready for

4 For a detailed discussion of computation and retrieval in language processing, see also Nooteboom, Weerman, and Wijnen (2002).

retrieval when a horse comes in focus. Kuiper's analyses show that the commentators strategically used strings of formulae that allowed them to keep talking fast and efficiently. The result is highly fluent, fast paced speech in which even occasional instances of cognitive disfluency are masked by the use of these formulae. Kuiper concludes from his analyses that the occurrence of formulaic sequences in spontaneous speech is a function of the pressure on the speaker's working memory and processing capacity. His findings also suggest that speakers can strategically choose formulaic language as a tool to maximize fluency in situations that require close attention and accuracy. The relationship between formulaic sequences and fluency is supported by a corpus analysis on speech pauses in spontaneous L1 speech (Erman, 2007). Pausing within formulae was rare, with pauses mostly occurring in open, non-prefabricated structures.

In the domain of L2 research, the available evidence is also scarce and mostly qualitative rather than quantitative. Dechert (1983) analyzed the spontaneous speech of a German learner of English and found that smooth and fluent production was related to the use of formulaic sequences, which he referred to as "islands of reliability". Similar results are reported by Wood (2006), who traced the development of formulaic language use in L2 learners of English across a period of six months. He concluded that learners used formulaic sequences to reduce hesitation and lengthen runs.

The findings from these L1 and L2 language production studies are generally in line with Pawley and Syder's (1983) hypothesis that the use of formulaic language is a key factor for the production of fluent speech, but to our knowledge there is no study that systematically and quantitatively analyzes the relationship between formulaic language and fluency in the spontaneous speech of bilingual and monolingual speakers.

5.1.4 Evidence for a smaller repertoire of formulaic sequences in L2 speakers

In their definition of the "puzzle of nativelike selection", Pawley and Syder (1983) state that in order to achieve native-like control, language learners must develop a way of distinguishing grammatical utterances that are native-like from those that are non-native-like, and that especially those learners who have to rely on grammar books, rather than on immersion in the speech community, have difficulties doing so.

The idea that non-native speakers use fewer formulaic sequences than native speakers and that their use deviates from native patterns is backed up by analyses

of L2 speech. Phraseological analyses based on written corpora, such as the *International Corpus of Learner English* (ICLE), show that advanced learners do use collocations in their L2, but often not in a native-like way: For example, Granger (1998) compared the use of collocations with intensifying adverbs (e.g., *perfectly normal*) in native English writing and that of advanced French-speaking learners of English. She concluded that language production in L2 speakers was characterized by both underuse and overuse of formulaic language, with overuse resulting from a limited repertoire of sequences that these speakers feel comfortable using. These findings are confirmed by more recent observations from Chen and Baker (2010) and Ädel and Erman (2012), who found similar patterns in an analysis of automatically extracted highly frequent sequences of words from a corpus of L1 and L2 academic English. Taken together, the results indicate that in L2 speakers, the repertoire of formulaic sequences is less well-defined than in native speakers. This seems to be driven by either a lack of representation (i.e., some sequences are not represented as such), or by an underspecification of a sequence's referential scope.

Studies of written corpora can be informative with respect to the underlying knowledge of collocations and other formulaic sequences that can be reproduced in the absence of constraints on time and working memory. However, these constraints are at the core of the argument made by Pawley and Syder (1983): Creating spoken utterances from scratch is an effortful task that may result in disfluencies, and the use of formulaic language might be a means to increase fluency, especially when cognitive resources are limited. It is therefore important to include data from spoken corpora into the analysis.

What do we know about the use of formulaic language in the spontaneous speech of L2 speakers? The available evidence, focusing on the relationship between proficiency and the use of collocations among L2 speakers, reaches mixed conclusions. In a corpus of conversations by advanced Swedish learners of French at three different levels of proficiency, Bartning, Forsberg and Hancock (2009) observed that only speakers at the lowest level of proficiency differed significantly from native speakers of French with respect to the number of lexical collocations. In contrast, Forsberg (2010) found that the average number of collocations rose significantly with the level of proficiency, with only very advanced (immersed) learners using them as frequently as native speakers. In her analyses, she distinguished between

formulaic sequences with different functions (referential, textual and communicative; see Granger & Paquot, 2008). Especially with respect to the number of referential sequences (e.g., *je m'appelle X* 'my name is X'), she concluded that their use is a very late or even native-like feature in the L2. The use of textual sequences (e.g., *un peu* 'a little bit'), however, did not vary with expertise. Forsberg argues that this is due to the fact that the number of textual sequences is much smaller than that of referential ones, and that textual sequences are likely to be learned early on in adult L2 acquisition.

In sum, the findings on collocations in spontaneous L2 speech are in line with the idea that *nativelike selection* is difficult to achieve: L2 speakers use collocations less frequently than native speakers and fluctuate between under- and overuse, dependent on the type of collocation. However, the relationship between L2 proficiency and the use of collocations is still underspecified. Götz (2013), who reported underuse of 3-grams and 4-grams among advanced German learners' spoken English, concluded that even these proficient speakers spend cognitive resources "on building their utterances from scratch instead of relying on highly frequent native-like chunks of words that are easily retrieved" (p. 106). She goes on to reason that not using formulaic sequences may be a key factor in explaining reduced fluency, even in advanced L2 speakers.

5.1.5 Summary and research questions

It has been hypothesized that L2 fluency is a function of the availability of formulaic sequences that can be retrieved from the mental lexicon with minimal computational effort (Pawley & Syder, 1983). The literature on the comprehension and production of formulaic language that we have reviewed is in line with this hypothesis, but the available evidence for a link between formulaic speech and fluency is mostly indirect.

The present study has been designed to provide a systematic quantitative analysis of L1 and L2 speech that directly targets the relationship between the use of formulaic language, fluency and proficiency. It expands our study on the relationship between proficiency and fluency (Bergmann et al., 2015) with an analysis of the formulaic language used in the same corpus of spontaneous speech. The corpus includes German movie retellings from a group of immersed L2 speakers of German with English as

their L1, a group of monolingual L1 speakers of German, and a group of German L1 attriters who at the time of testing had been residents of the United States of America or Canada for at least five years. To identify and classify formulaic sequences, we have selected the taxonomy of Granger and Paquot (2008), which has been designed to cover the full spectrum of literal and figurative formulaic sequences.

Based on these classifications, we want to answer three main research questions.

(1) Are there quantitative differences between monolingual and bilingual speakers with respect to the use of formulaic sequences in our corpus of movie retellings?
(2) Is there a clear relationship between the number of formulaic sequences that speakers use in spontaneous speech and other linguistic variables (such as global linguistic proficiency or lexical diversity)?
(3) Is the number of formulaic sequences related to the degree of speech fluency?

Our approach to answering these questions differs from earlier research in two important ways. First, we aim at classifying all types of formulaic sequences in the corpus, not only collocations, as speakers with different levels of proficiency may not only differ with respect to the number of formulaic sequences, but also with respect to the types they prefer. Second, we include a group of L1 attriters into the comparison. They are native speakers of the testing language (German) and had fully acquired it before emigrating to the US, but in the meantime have become highly proficient speakers of another language (English). They are therefore comparable to the monolingual control group with respect to their acquisition of German and comparable to the L2 speakers with respect to a possible role of cross-language interference and competition. By including L1 attriters into our analyses, we avoid the confound of incomplete acquisition and language competition that is inherent to analyses of L2 speakers' data.

What are our expectations with respect to the use of formulaic sequences? Following both Pawley and Syder's (1983) hypotheses and the observations in the literature reported above, we expect a clear difference between the monolingual controls and the L2 speakers. If formulaic sequences are indeed difficult to acquire in an L2, we should find that overall L2 speakers use them significantly less frequently than monolingual native speakers. This effect may be further enhanced by effects of language competition, that is, L2 speakers may actively need to inhibit formulaic

sequences from their L1. The same holds for the L1 attriters. Whether they pattern with the controls or with the L2 speakers, however, is an open question, as the use of formulaic language may be affected by language competition in at least two ways: Attriters may, for example, find it more difficult to access L1 sequences. Alternatively, they may resort to the use of formulaic speech in order to reduce the processing burden, similar to Kuiper's horse race commentators.

With respect to the relationship between formulaic sequences and fluency, our predictions are again driven by Pawley and Syder's (1983) hypothesis that the retrieval of these sequences from the mental lexicon enhances speech fluency by means of reducing the processing costs related to speech planning. This means that we predict a negative correlation between the number of formulaic sequences and the number of disfluencies per speaker. This effect should hold for all three groups of speakers in our corpus, but may be strongest in the two bilingual groups.

5.2 Method

5.2.1 Participants

Three groups of 20 participants each were included in the study: (1) L1 speakers of German who were living in Germany, used no L2 in their daily lives and were tested in Germany (= controls); (2) L1 speakers of German who were living in the United States of America, used English as their L2 and were tested in the US (= L1 attriters); (3) L1 speakers of English who were living in Germany, used German as their L2 and were tested in Germany (= L2 speakers). The participant characteristics are summarized in Table 10.

There were small demographic differences between the three groups of speakers: First, there were significantly fewer males among the monolingual controls than in the L2 speakers (Kruskal–Wallis test: $W = 140$, $p = .043$); the L1 attriters differed neither from the controls ($W = 190$, $p = .696$) nor from the L2 speakers ($W = 150$, $p = .099$). Second, the L1 attriters were significantly younger than the controls ($W = 290$, $p = .015$), but not than the L2 speakers ($W = 177$, $p = .542$), and the latter did not differ significantly from the controls ($W = 251$, $p = .171$). Third, the significant group differences with respect to education were due to the controls having received less formal education than the L1 attriters ($W = 102$, $p = .005$) and L2 speakers ($W = 91$, $p = .002$), who did not differ significantly from one another ($W = 187$, $p = .692$).

Table 10: Summary of participant characteristics.

	Control group (n = 20)	*L1 attriters (n = 20)*	*L2 speakers (n = 20)*	*Group comparison*[a]
		Mean (SD; Range)		
Male	**15 %**	**20 %**	**45 %**	$\chi^2 = 5.196$, $p = .074$
Age	**48.6** (5.8; 39–58)	**42.6** (9.2; 29–62)	**45.3** (12.0; 25–73)	$\chi^2 = 5.4923$, $p = .064$
Education[b]	**2.6** (1.0; 1–4)	**3.5** (0.7; 2–4)	**3.6** (0.6; 2–4)	$\chi^2 = 12.5615$, $\mathbf{p = .002}$
Age of emigration to L2 setting	–	**28.0** (5.3; 23–40)	**25.1** (4.1; 20–3)	W = 263, $p = .089$
Length of residence in L2 setting	–	**14.5** (7.9; 7–34)	**26.1** (11.5; 7–53)	W = 71, $\mathbf{p < .001}$

a: Kruskal–Wallis or Wilcoxon rank-sum test.
b: Education was coded on the following scale: 1 = lower secondary education (German 'Volksschule/ Hauptschule' or equivalents); 2 = intermediate secondary education (German 'Realschule' or equivalents); 3 = higher secondary education (German '(Fach-)Abitur' or equivalents); 4 = university degree.

5.2.2 Materials and procedure

Participants watched a ten-minute excerpt from the silent movie *Modern Times* (1936), starring Charlie Chaplin and Paulette Goddard. It runs from a scene showing Charlie Chaplin's failed attempt at working at a shipyard (about 33 min into the film) to a scene in which a police officer chases the main characters from a lawn. This sequence has been used in L2 research for some twenty years (e.g., Perdue, 1993; Schmid & Köpke, 2007) to elicit longer stretches of spontaneous speech.

Immediately after the end of the screening, participants were asked to retell in German what they had seen. The experimenter prompted participants to start their retellings by using a global question such as *Was haben Sie gesehen?* 'What did you see?' and refrained from interrupting participants during their retellings to avoid influencing their linguistic behavior. If participants failed to mention important parts of the film, the experimenter reminded them of the scenes at the end of their

retelling by referring to scenes that preceded or followed the omitted scene and asking how the story proceeded.

After the retelling, participants completed a pen-and-paper proficiency test (C-test). Each participant was tested individually by a native speaker of German in two-hour testing sessions that included a variety of language and memory tasks. The spontaneous speech task was not the first in the testing session, which gave participants some time to get into the language mode of the experimental language (Grosjean, 1982, 1998). The retellings were taped using different types of digital audio recorders and microphones, all producing 16-bit WAV files with a sampling frequency of 44,100 Hz.

5.2.3 Data selection, transcription and coding

The recordings were filtered for noise and their volume was normalized prior to transcription. In most cases, participants produced a continuous and uninterrupted retelling that was analyzed in its entirety. In a minority of cases, reminding participants of scenes they had left out led to another longer speech stretch, so these data were included in the analysis as well. However, we discarded all answers to the experimenter's questions that were shorter than three consecutive sentences. All data were transcribed by native speakers of German following the standards of the CHAT transcription format and were processed using the CLAN software (MacWhinney, 2014a, 2014b).

The classification of sentence elements as formulaic sequences was based on the categories defined by Granger and Paquot (2008); their categories are exhaustively listed in Table 11 and will be discussed below. The coding of the use of formulaic sequences was done by the first and second author, both native speakers of German. In order to standardize the implementation of the classification system, sample transcriptions were processed by both coders. Comparisons revealed few, if any, differences. Remaining disagreements were discussed among the coders and resolved. Unclear cases that were encountered during the coding of the actual corpus were treated likewise. Spot checks of completeness and accuracy that were done on the coded transcripts revealed few problematic decisions.

5.2.4 Classification of formulaic sequences

We selected the system developed by Granger and Paquot (2008) because of its clarity and the large variety of types covered. The authors distinguish eighteen different

Table 11: Classification of formulaic sequences by Granger and Paquot (2008).

Formulaic sequences		
Referential function	*Textual function*	*Communicative function*
› Lexical collocations *(sharp braking, closely linked)* › Grammatical collocations *(interested in, depend on)* › Idioms *(to spill the beans)* › Irreversible bi- and trinomials *(bed and breakfast, kith and kin)* › Similes *(as old as the hills)* › Compounds *(goldfish, blow-dry)* › Phrasal verbs *(blow up, make out)*	› Complex prepositions *(apart from, irrespective of)* › Complex conjunctions *(even though, given that)* › Linking adverbials *(in other words, what is more)* › Textual sentence stems *(another thing is, I will discuss)*	› Speech act formulae *(Good morning! Take care!)* › Attitudinal formulae *(in fact, to be honest)* › Proverbs and proverb fragments *(A bird in the hand ...)* › Commonplaces *(Enough is enough)* › Slogans *(Make love, not war)* › Idiomatic sentences* › Quotations* * No examples given in the original classification

categories, ranging from lexical collocations to quotations, with the intention to cover a broad spectrum of formulaic language used in written and spoken language. These categories are listed in Table 11. For detailed definitions of all categories, we refer the reader to Granger and Paquot (2008, pp. 22–24). Here, we focus on the definition of collocations, as these were the most relevant type of formulaic sequence in our corpus. Examples below are taken from our own corpus; italics denote words that belong to the sequence in question.

Granger and Paquot (2008) distinguish between lexical collocations and grammatical collocations. Lexical collocations are defined as conventionalized combinations of content words, such as in (1):

(1) Es kommt zu (ei)ner *scharfen Bremsung*.
It comes to a *sharp braking*.
'There is sharp braking.'

The classification of lexical collocations was based on the criterion that at least two content words had to be involved, either with or without a function word. An utterance as in (2), involving an indefinite determiner, has also been classified as a lexical collocation.

(2) Die junge Frau *macht ein Gesicht*.
The young lady *makes a face*.
'The young lady pulls a face.'

The most common lexical collocations concern combinations of adjectives and nouns, as shown in (1), combinations of nouns and verbs, as shown in (2), and combinations of adjectives and verbs, as in (3):

(3) Dann ist ihm die Sache *zu heiß geworden*.
Then is him the thing *too hot become*.
'Then things got too hot to handle for him.'

A subtype of lexical collocations is the category of irreversible bi- and trinomials. This refers to sequences of two or three content words the order of which is fixed. Only few formulaic sequences qualify for inclusion in this category, such as the one in (4):

(4) *Schließlich und endlich* haben die beiden sich auf den Weg gemacht.
Ultimately and finally have the two themselves on the way made.
'When all was said and done, the two of them hit the road.'

Conventionalized combinations that involve no more than one content word and one or more function words were classified as grammatical collocations, shown in (5):

(5) Es sah so aus, als wären die ganz *interessiert an*einander.
It looked as were they fairly *interested at* another.
'It seemed as if they were fairly interested in one another.'

The relevant sequence here concerns the combination of *interessiert* 'interested' with *an* 'at', the only possible preposition to specify the object of interest in German.

Two categories from Granger and Paquot's original system were removed from our own classification: The classification of 'compounds' and 'phrasal verbs' cannot be directly transferred from English to German. A useful distinction of degrees

of lexicalization in German compounds would have to go beyond semantics and include, among other aspects, morphological complexity. Also, it is unclear which types of compounds and phrasal verbs have to be considered part of formulaic language in the narrow sense. As these distinctions are beyond the scope of our study, we excluded these categories from our analyses.

5.2.5 Additional variables

We collected data on three variables unrelated to the use of formulaic sequences that reflect different aspects of proficiency and language use in our participants.

First, general linguistic proficiency was assessed using a pen-and-paper C-test. Participants filled in two German texts, constructed by Schmid (2011), in which parts of every other word were missing. The share of incomplete words in the two texts was 37 or 41%, respectively. Answers were coded on a scale from 0 (no response provided) to 9 (same word as in original text); spelling errors were not counted as incorrect (for the full coding system, see Schmid, 2016).

Second, lexical diversity in the retellings was estimated using the variable D, which involves random samplings of tokens from the transcript under investigation. This avoids the sensitivity to sample size, which is characteristic of other measures of lexical diversity, such as type/token ratio (McCarthy & Jarvis, 2010). We calculated D using the VOCD routine provided by the CLAN program (MacWhinney, 2014b).

Third, the frequency of occurrence of modal particles in the retellings was counted. Modal particles in German are uninflected words that are used to express speaker attitudes, mainly in spoken language. They are often homonymous with adjectives, adverbs or conjunctions, and their exact meaning is often difficult to pinpoint (Burkhardt, 1994). These factors make particles hard to acquire for L2 speakers (Möllering, 2001; Möllering & Nunan, 1995), which is why they have been used as a proxy of general pragmatic competence in spoken language (Belz & Vyatkina, 2005). In the coding process, potential modal particles were identified automatically in the transcripts, based on a list of all words that can be used as modal particles, as labeled in the electronic edition of a German dictionary (Dudenredaktion, 2011). It was then decided on a case-by-case basis whether the word was actually used as a modal particle in the given context. The coding of the modal particles was done by the second author, who is a native speaker of German. Few difficult cases were

Table 12: Length of retellings (number of tokens per participant).

	Average (Range)	*Median*	*SD*
Monolingual controls	**791** (318–2311)	706	450
Bilingual L1 attriters	**1220** (667–2184)	1218	400
Bilingual L2 speakers	**754** (272–1394)	704	322

encountered. In these cases, the recording of the retelling was consulted, as the pronunciation usually reveals the intended meaning.

5.3 Results

5.3.1 General sample characteristics

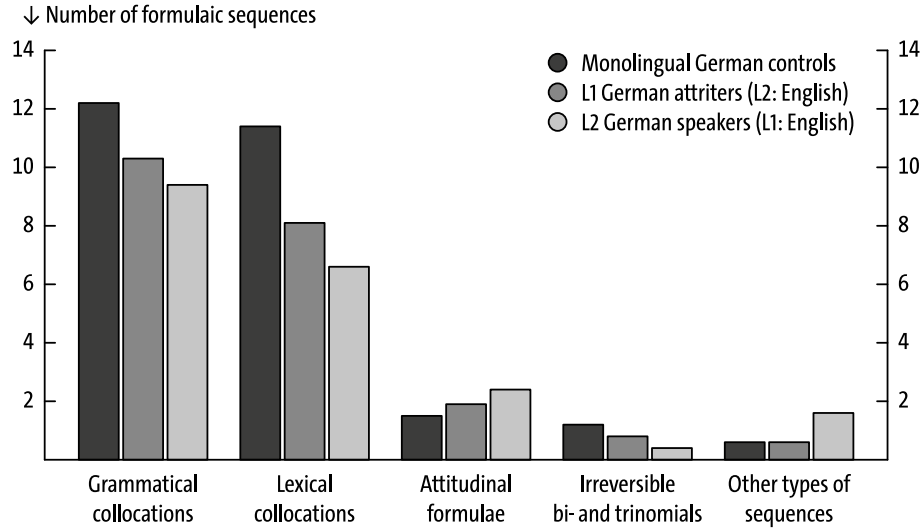
All sixty speakers produced coherent German retellings of sufficient length. Table 12 shows that, on average, the attriters produced longer retellings than the monolingual controls and the L2 speakers, who did not differ much from each other. This pattern was confirmed by a significant Kruskal–Wallis test of the group differences, followed by three Wilcoxon rank-sum tests (Kruskal–Wallis: $\chi^2 = 15.8671$, $p < .001$; Wilcoxon: Controls vs. L1 attriters, $W = 73$, $p < .001$; controls vs. L2 speakers, $W = 197.5$, $p = .957$; L1 attriters vs. L2 speakers, $W = 327$, $p < .001$).

In the remainder of this section, all comparisons of group averages will be based on this combination of nonparametric tests. To account for the number of comparisons, the significance level for the pairwise comparisons has been adjusted to .05/3 = .017.

5.3.2 Use of formulaic sequences

Figure 6 shows the average frequency counts for the various types of formulaic sequences in the three groups of speakers. The four types of sequences that appeared at least once per 1000 words across the corpus (grammatical collocations; lexical

Figure 6: Average counts of different types of formulaic sequences (per 1000 words; following Granger & Paquot, 2008) in the retellings.



collocations; attitudinal formulae; and irreversible bi- and trinominals) are represented by separate sets of bars. The remaining categories have been summarized in the category 'Other'.

There are three interesting observations. First, each type of formulaic sequence was used by speakers from each group. Second, with the exception of Other, the rank order of the categories was identical in all groups: Grammatical collocations were most frequent, followed closely by lexical collocations, then by attitudinal formulae and binomials. Third, when comparing the groups, we see a steady decrease in the counts of grammatical and lexical collocations from monolingual controls to L1 attriters and L2 speakers. For the lexical collocations, this decrease is steeper than for the grammatical collocations. In the low-frequency categories, the differences between groups are less pronounced. In the category 'Attitudinal formulae' (e.g., *I think that*), we see the opposite pattern.

More detailed descriptives and the results of the inferential tests of these differences, given in Table 13, show a significant effect of speaker group only for the lexical collocations.

Table 13: Frequency of formulaic sequences (per 1000 words).

	Monolingual controls	*Bilingual L1 attriters*	*Bilingual L2 speakers*	*Group comparison*[a]
Category		*Mean (SD; Range)*		
Grammatical collocations	**12.2** (5.6; 2.4–25.2)	**10.3** (4.0; 3.1–20.4)	**9.4** (5.5; 0–23.2)	$\chi^2 = 3.499$, $p = .174$
Lexical collocations	**11.4** (6.3; 1.8–28.4)	**8.1** (4.0; 1.9–15.7)	**6.6** (2.4; 2.7–11.5)	$\chi^2 = 9.148$, $\mathbf{p = .010}$
Attitudinal formulae	**1.5** (1.9; 0–6.5)	**1.9** (1.6; 0–5.6)	**2.4** (3.0; 0–9.8)	$\chi^2 = 0.942$, $p = .625$
Irreversible bi- and trinomials	**1.2** (1.7; 0–6.3)	**0.8** (1.2; 0–3.9)	**0.4** (1.0; 0–4.1)	$\chi^2 = 5.331$, $p = .070$
Other formulaic sequences	**0.6** (1.0; 0–3.1)	**0.6** (0.8; 0–2.4)	**1.6** (3.8; 0–16.8)	$\chi^2 = 2.889$, $p = .236$

a: Kruskal–Wallis test

Pairwise comparisons of these averages further reveal that only the largest group difference – between controls and L2 speakers – is significant, with the L2 speakers using fewer lexical collocations than the monolingual controls (Controls vs. L2 speakers: $W = 313$, $p = .002$). The L1 attriters, who descriptively find themselves in between the controls and the L2 speakers, did not differ significantly from either group (Controls vs. attriters: $W = 267$, $p = .072$; L1 attriters vs. L2 speakers: $W = 237$, $p = 324$).

Summing up, the results show a highly similar pattern of formulaic language use in the three groups of speakers, with grammatical and lexical collocations standing out as the most frequently used categories in all three groups. However, the significant decrease in the number of lexical collocations used by non-native speakers also indicates that native and non-native speakers use formulaic language differently.

5.3.3 Additional linguistic variables

The results of the C-test, the D score and the number of modal particles are summarized in Tables 14 (descriptives) and 15 (inferential statistics). D and the number of particles have been derived from the retelling.

Table 14: Additional linguistic variables.

		Average (Range)	*Median*	*SD*
C-test[a] *(General proficiency)*	*Monolingual controls*	**82 %** (71 – 89 %)	83 %	5 %
	Bilingual L1 attriters	**95 %** (79 – 99 %)	96 %	4 %
	Bilingual L2 speakers	**86 %** (66 – 99 %)	87 %	9 %
D score (Lexical diversity)	*Monolingual controls*	**75** (47 – 129)	68	23
	Bilingual L1 attriters	**66** (41 – 80)	66	11
	Bilingual L2 speakers	**64** (23 – 107)	63	21
Number of modal particles[b]	*Monolingual controls*	**24** (6 – 49)	23	12
	Bilingual L1 attriters	**17** (2 – 42)	16	11
	Bilingual L2 speakers	**13** (0 – 38)	12	10

a: Percentage of correct answers; **b:** Per 1000 words

Table 15: Group differences in the three additional variables.

			Wilcoxon rank-sum tests					
	Kruskal–Wallis omnibus test		*Control vs. L1 attriter*		*Control vs. L2 speaker*		*L1 attriter vs. L2 speaker*	
	χ^2	*p*	*W*	*p*	*W*	*p*	*W*	*p*
C-test	27.5361	**< .001**	15	**< .001**	139	.101	334	**< .001**
D score	1.3862	.500	–	–	–	–	–	–
Particles	8.2871	**.016**	266	.076	304	**.005**	241	.273

The results of the C-test show that all three groups were highly proficient with respect to their declarative knowledge of German vocabulary and morphosyntax. The significant group differences were due to the attriters, who scored significantly higher than both the monolingual controls and the L2 speakers. The latter did not differ from each other.

The D score is a measure of lexical diversity. Although the monolingual control speakers' scores were slightly higher than those of the other two groups, the difference was not significant, confirming the bilinguals' high proficiency.

As a proxy variable for pragmatic competence in spoken language, we included the number of modal particles per 1000 words. We find that the control group used almost twice as many particles as the L2 speakers and that the attriters were placed in between these two groups. While the difference between the control group and the L2 speakers was significant, the attriters did not differ significantly from either of the two groups.

Taken together, these measures indicate that possible differences in how the speakers use formulaic sequences, which is the main focus of this article, cannot be attributed to a low level of overall proficiency or a lack of vocabulary on the single-word level in the L2 speakers. However, the significant differences with respect to particle use illustrate that, in multilingual speakers, "native-like competence" is a multifaceted phenomenon.

5.3.4 Relationship between the use of formulaic language and additional variables

Our counts have shown that grammatical and lexical collocations are by far the most frequent types of formulaic sequences in our corpus of spontaneous speech data. We performed correlational analyses to reveal possible relationships between these categories and additional linguistic variables.

Table 16 shows the correlations between the use of collocations and the results of the C-test, the D score and the number of modal particles that the speakers used. Scatter plots of selected correlations are provided in Figure 7.

There were moderately strong and reliable positive correlations between the number of grammatical collocations and the results of the C-test in the controls and L2 speakers. In light of the C-test results presented above, it seems possible that the lack of a correlation in the attriters is due to the fact that their C-test scores were at ceiling and showed little variance. We therefore conclude that the use of

Figure 7: Correlations of the C-test score with the number of grammatical collocations (top row) and of the D score with the number of lexical collocations (bottom row).

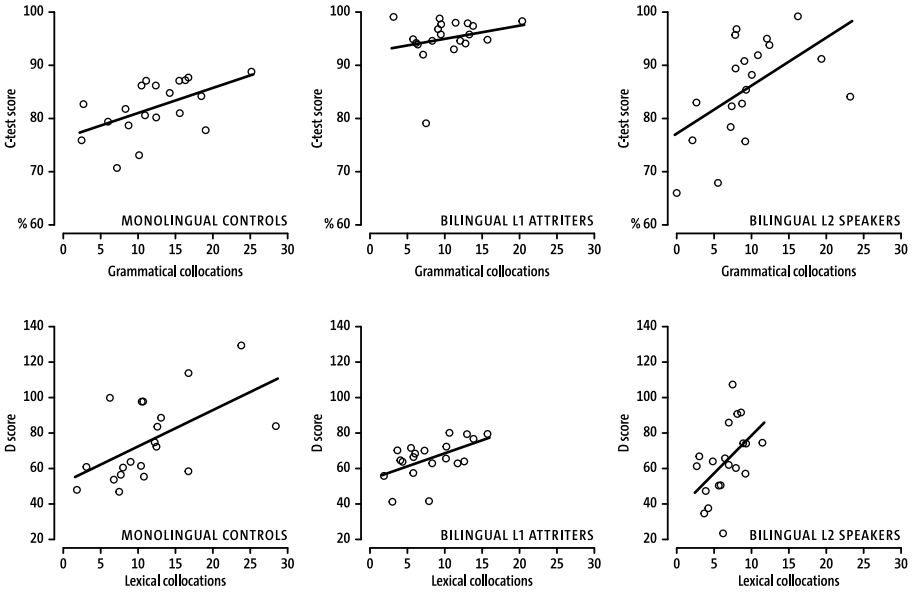


Table 16: Correlations between grammatical collocations (GC) or lexical collocations (LC) and additional linguistic variables.

		Monolingual controls		*Bilingual L1 attriters*		*Bilingual L2 speakers*	
		rho	*p*	*rho*	*p*	*rho*	*p*
C-test (General proficiency)	GC	**.55**	**.012**	.24	.300	**.61**	**.005**
	LC	.19	.432	-.08	.753	−.22	.348
D score (Lexical diversity)	GC	.22	.346	.44	.846	.22	349
	LC	**.55**	**.011**	**.52**	**.018**	**.55**	**.011**
Number of particles	GC	−.20	.398	.31	.191	.42	.068
	LC	−.40	.082	.10	.663	-.03	.907

grammatical collocations and the C-test reflect similar aspects of language proficiency. In contrast, the remaining two measures do not show significant correlations with the grammatical collocations.

With respect to lexical collocations, we found reliable positive correlations with the D score, a measure of lexical diversity, in all three groups of speakers. We conclude that, to some extent, these two measures reflect similar types of knowledge or abilities. No other correlations with lexical collocations reached significance.

In sum, the results support the idea that grammatical and lexical collocations are distinct phenomena, reflecting different aspects of language proficiency. Interestingly, the use of particles seems to reflect yet another dimension of linguistic competence, as it is related to the use of neither grammatical nor lexical collocations.

5.3.5 Relationship between the use of formulaic language and speech fluency

Table 17 shows the correlations between the number of collocations and four fluency measures: speech rate; the number of pauses; the number of repetitions; and the number of self-corrections. The fluency measures are based on the same retellings that were also used in this study. The method and results of the fluency analyses are reported in detail in Bergmann et al. (2015). Scatter plots of selected correlations are provided in Figure 8.

The results show that only two fluency measures vary systematically with the use of collocations, and that these effects are almost exclusively reserved for lexical collocations. The exception to this finding is an unpredicted, but marginally significant positive correlation between the number of repetitions and the number of grammatical collocations in the L2 speakers. Most importantly, however, we find a strong and reliable negative relationship between the lexical collocations and two major fluency indicators in the L1 attriters and L2 speakers: the more lexical collocations these speakers produced, the fewer repetitions and self-corrections occurred. The relationship is somewhat less strong in the monolingual controls, where the correlation with self-corrections and repetitions is only marginally significant (assuming $\alpha = .017$).

5.4 Discussion

Native speakers and listeners have access to a large repertoire of formulaic sequences, such as *How do you do?*, which can readily be retrieved from the mental lexicon

Figure 8: Correlations of the number of lexical collocations with the number of repetitions (top row) and the number of self-corrections (bottom row).

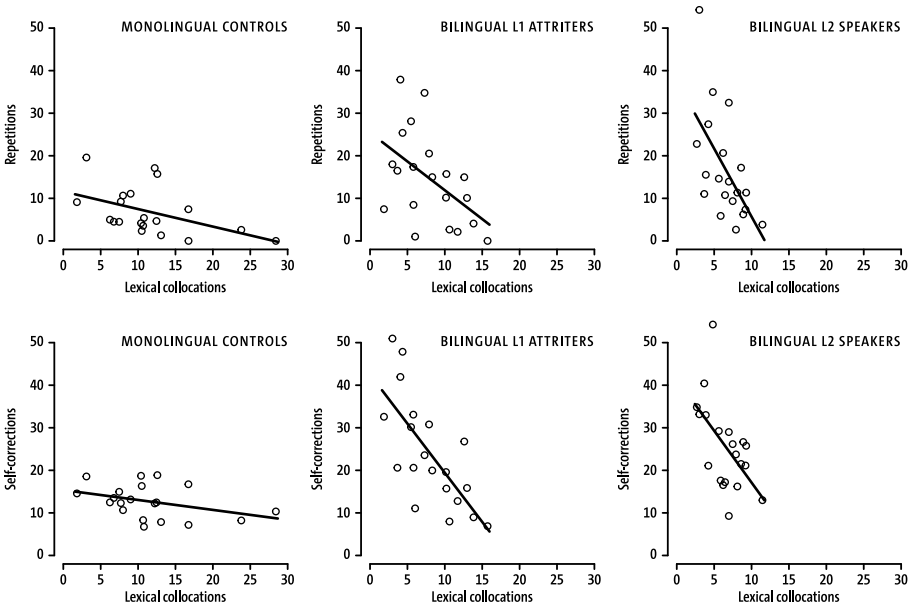


Table 17: Correlations between grammatical collocations (GC) or lexical collocations (LC) and fluency measures.

		Monolingual controls		*Bilingual L1 attriters*		*Bilingual L2 speakers*	
		rho	*p*	*rho*	*p*	*rho*	*p*
Speech rate	*GC*	.01	.970	.08	.748	−.09	.705
	LC	−.04	.853	−.21	.373	.30	.207
N of pauses	*GC*	.20	.393	−.09	.705	.23	.321
	LC	.04	.871	.14	.530	−.26	.251
N of repetitions	*GC*	.03	.885	−.21	.384	.48	.031
	LC	−.47	.036	**−.55**	**.012**	**−.59**	**.006**
N of self-corrections	*GC*	.22	.359	−.39	.091	−.08	.734
	LC	−.43	.059	**−.77**	**< .001**	**−.55**	**.012**

(Wray, 1999, 2002). Pawley and Syder (1983) hypothesized that the use of this type of formulaic language not only reflects native-like proficiency, but also enhances speech fluency. The ability to retrieve large chunks of information from long-term memory is assumed to free the speaker from the limitations of their processing system, thereby reducing the number of errors and error corrections.

To test this hypothesis, we analyzed the use of formulaic language and its relation to speech fluency in the spontaneous speech of L1 and L2 speakers of German. We specifically wanted to know whether there are quantitative differences between monolingual and bilingual speakers with respect to the use of formulaic sequences, how the number of formulaic sequences that speakers use in spontaneous speech relates to other linguistic variables, such as general proficiency, and whether the number of formulaic sequences that speakers use is related to the degree of speech fluency.

We compared highly proficient immersed L2 speakers of German (L1: English) to two groups of native speakers, using a spontaneous speech corpus based on movie retellings. First, we chose a group of monolingual speakers of German in order to establish a native baseline. Second, we analyzed the speech of a group of bilingual German L1 attriters who lived in the United States and had been immersed in their L2 (English) for more than 14 years on average. In an earlier analysis of this corpus with respect to speech fluency only, we have shown that the L1 attriters and the L2 speakers are more comparable to one another when it comes to effects of competition between their L1 and L2 (Bergmann et al., 2015). The present analysis focused on the number and types of formulaic sequences, using a classification system defined by Granger and Paquot (2008).

Our analyses show that all three groups of speakers used formulaic language in their retellings, with the two most common types of sequences being lexical and grammatical collocations. All other types of formulaic sequences only very rarely occurred in our sample. A statistical comparison revealed that the three groups only differed with respect to the number of lexical collocations: The L2 speakers used significantly fewer lexical collocations than the monolingual speakers, while the attriter group differed from neither of them. The pattern of results indicates that the frequency with which speakers use lexical collocations in spontaneous speech is a function of both proficiency and language competition. We therefore investigated the relationship between three other linguistic measures – related to global

proficiency, lexical diversity and the use of modal particles – and the use of collocations in our corpus.

First, a comparison of the groups with respect to their scores on a global proficiency test (C-test) confirmed that all three groups were highly proficient, with the attriters scoring highest on the test. Moreover, we found that the higher the speakers' global proficiency, the more grammatical collocations they used. This relationship is strong and reliable among the monolingual controls and the L2 speakers. In the attriters, we descriptively see a trend in the same direction. Interestingly, we did not find a systematic relationship between the number of lexical collocations and the C-test score.

Second, we determined the level of lexical variability in the retellings, which by itself did not distinguish between the groups, and its relationship with the use of collocations. We found a strong and reliable positive relationship between D and the number of lexical collocations in all three groups: the higher the degree of lexical diversity on the single-word level, the more lexical collocations the speakers produced. However, mirroring our findings for the C-test, we found no such relationship for grammatical collocations.

Third, as an additional measure of linguistic competence in spoken language, we assessed the number of modal particles in the retellings. The use of modal particles is a typical feature of native German, but it is difficult to acquire for late learners of German (e.g., Möllering & Nunan, 1995). Indeed, we found that the control group used almost twice as many particles as the L2 speakers and that the attriters were placed in between these two groups. While the difference between the controls and the L2 speakers was significant, the attriters did not differ significantly from the other groups. The result shows that, even though the L2 speakers were highly proficient in German, they were not native-like in all aspects of language use. At the same time, the intermediate position of the attriters indicates that competition from another language can result in non-native-like performance and that part of the effect in the L2 speakers can be attributed to this factor. However, when correlating the number of particles to the number of collocations, we did not find a systematic relationship in any of the three groups. This suggests that the pragmatic ability to use modal particles relates to a different type of proficiency than the use of collocations.

In sum, this pattern of correlations suggests that the use of lexical and grammatical collocations reflects different aspects of language proficiency. While the

use of grammatical collocations relates to declarative knowledge and global proficiency (including morphosyntax), the use of lexical collocations is linked to lexical diversity, as deduced from online speech performance. Neither type of collocation is related to the use of modal particles, which reflects high-level pragmatic abilities in German.

In a second set of analyses, we turned to the core question of the present study: Does the use of formulaic language enhance a speaker's fluency? To answer this question, we correlated the numbers of grammatical and lexical collocations with four measures of speech fluency. Only lexical collocations showed strong and reliable negative correlations with two measures of disfluency (number of repetitions and number of self-repairs) in the two bilingual groups: the more lexical collocations attriters and L2 speakers used, the more fluently they spoke. For the monolingual controls, the relationship was only marginally significant when corrected for multiple comparisons, even though they use lexical collocations most frequently. This may reflect a ceiling effect, as the monolinguals were the most fluent of the three groups.

Especially the latter issue confirms our choice to include L1 attriters as a bilingual control group, as they allow us to investigate the relationship between the use of collocations and native speech fluency more closely. Attriters produce natural speech in a situation that puts their resources under pressure. Like the L2 speakers, they must invest part of their processing capacities into language selection and speech monitoring. As a consequence, their fluency performance is not at ceiling. However, the resulting decrease in fluency means that there is also more room for improvement. It is in line with this argument that we found a significant correlation between fluency and the use of lexical collocations among both the L1 attriters and the L2 speakers.

Taken together, our findings strongly support Pawley and Syder's (1983) hypothesis about the relationship between the use of formulaic language and native-like fluency. At the same time, we find that not all types of formulaic language are alike: Although there is a clear relationship between global proficiency and grammatical collocations, the use of this type of formulaic sequence is not related to fluency in our speaker groups.

What are the cognitive mechanisms that are responsible for the fluency effect? We would like to argue that the advantage arises from the way in which formulaic sequences are stored and accessed in the mental lexicon. Concrete ideas about the nature of representations involved and the way in which they are integrated in the

mental lexicon have been put forward in the psycholinguistic models of idiom production by Cutting and Bock (1997) and by Sprenger, Levelt and Kempen (2006). While disagreeing about the exact nature of the syntactic information that is stored with an idiom and the specific way in which single lexical items become involved, both models assume that idioms have their own entries in the mental lexicon. During speech production, the selection of these representations eventually results in the retrieval of pre-specified clusters of single lexical items and their syntactic arrangements. When efforts for lexical retrieval, syntactic encoding and linearization are reduced substantially, possible sources of errors are minimized and additional resources for speech monitoring become available. As a result, the models predict that speakers are able to produce relatively large stretches of speech without the need for pauses, repetitions or self-corrections. In addition, due to the incremental nature of language production (Kempen & Hoenkamp, 1987), more resources become available for formulating the speech that follows the pre-specified chunk.

How do these models of L1 idiom production apply to the production of formulaic language in L2 speakers? First, the models' scope can easily be expanded to non-figurative formulaic language, because the individual lexical items are not selected on the basis of their meaning. Instead, the production process is driven by the selection of a conceptual entry (lexical concept node) that ties the meaning of the fixed phrase itself to the remainder of the lexicon. In both approaches, the individual lexical entries that are selected for production may or may not be semantically related to the overall meaning of the phrase. Second, in the most parsimonious model for L2 production, the architecture of the language production system and the tools at its disposal are the same as those for the L1 (e.g., de Bot, 1992). Consequently, possible differences between L1 and L2 performance are best explained by quantitative and qualitative differences in the underlying representations. L2 speakers know fewer formulaic sequences in their L2 than native speakers do, the ones they do know are less well represented, and L1 representations may compete for selection. In terms of a processing model, this will result in longer processing times or even failure of retrieval, and ultimately in the need to generate new lexical strings instead. With limited resources being mostly directed at planning and reformulation, rather than, for example, speech monitoring, speech output will become disfluent.

In summary, we argue that the line of argument proposed by Pawley and Syder (1983) is fully congruent with current models of idiom processing. These models allow

us to pinpoint the potential source of the processing advantage that arises from the use of formulaic language. Specifically, they predict online formulation processes to be replaced with direct retrieval steps (e.g., of a superlemma). The gain of this type of processing 'shortcut' is twofold: First, direct retrieval entails that fewer errors arise during formulation. Second, more processing capacity becomes available that can be used for speech monitoring. Consequently, errors in phonological encoding and articulation can be caught faster and with higher precision, and formulation problems with respect to upcoming increments can be corrected early on. The application of psycholinguistic models of idiom production therefore predicts that the fluency gain that arises from formulaic language will concentrate largely on the formulaic sequences themselves as well as on the increments that directly follow them.

Future research will need to address a number of open questions. First, it is as yet unclear whether variations in semantic transparency, figurativeness, syntactic fixedness etc. are relevant dimensions with respect to the production of formulaic language. When it comes to fluency, does it matter which type of sequence the speaker uses? In order to answer this question, we will need a speech corpus that covers a large variety of communicative situations. More importantly, we need more fine-grained hypotheses, both about how the representations for various types of formulaic language differ from each other, and about how these differences affect processing and processing effort. Second, it would be interesting to analyze the point in an utterance at which disfluencies arise. In particular, it would be worthwhile to test the hypothesis that the segments that follow a formulaic sequence benefit most from the processing advantages that arise from the direct retrieval of speech chunks from the lexicon.

5.5 Conclusion

Considering our findings, and in line with Pawley and Syder (1983), we conclude that not having easy access to pre-specified chunks of formulaic language, as we have in our L1, is directly related to a lack of L2 fluency. In addition, the chunks that we prefer to use in our L1 possibly hamper access to formulaic sequences in our L2, and vice versa. When speaking the other language, we therefore not only find ourselves lost for words, but lost for chunks. As speaking in chunks is what makes us fluent, learning the equivalent of *How do you do?* is an excellent way to get started in a new language.